SENA
MINISTERIO DE TRABAJO Y SEGURIDAD SOCIAL
SERVICIO NACIONAL DE APRENDIZAJE
REGIONAL BOGOTA, CUNDINAMARCA
CENTRO DE ADMINISTRACION

IMPORTANTE.

Este es Materialen Prueba, sujeto a modificaciones. Prohibido su reproducción total o parcial.

UNIDAD ESTADIGRAFOS DE DISPERSION

SECTOR COMERCIO Y SERVICIOS

FAMILIA OCUPACIONAL TECNICO ADMINISTRATIVO

MODULO OCUPACIONAL INVESTIGACION DE MERCADOS

MODULO INSTRUCCIONAL INVESTIGACION MERCADOLOGICA

CODIGO: 25.01.12

ELABORADA POR: DOCTOR JORGE ELIECER CASTRO ROYETT

ASESORES METODOLOGICOS GUTNAR GOMEZ
MARIO PEÑARANDA

ILUSTRACIONES

C O N T E N I D O

INTRODUCCION

En la unidad que acaba de estudiar aprendió usted a manejar los estadígrafos conocidos con el nombre de Medidas de Tendencia Central, herramientas muy útiles para conocer bajo otro ángulo diferente al de las distribuciones de frecuencias vistas en las unidades No. 1 y No. 2, el comportamiento de cualquier serie estadística.

No obstante, la aplicación de tales medidas no es suficiente para el análisis de las características de una población. Se necesita de alguna medida adicional, adecuada para mostrar cómo se distribuyen o se dispersan los datos alrededor del promedio seleccionado. Tales medidas son las que estudiará en esta unidad bajo el nombre de Estadígrafos de Dispersión y constituyen otros instrumentos de uso frecuente entre quienes manejan la información estadística de las empresas o desarrollan investigaciones mercadológicas.

En consecuencia, estamos convencidos de su interés en aprender a usar correctamente estos estadígrafos como lo hizo con los anteriores, razón por la cual le deseamos éxito en su empeño.

OBJETIVO TERMINAL

Dada la clasificación y tabulación de los datos, al terminar la unidad, el estudiante estará en capacidad de calcular los estadígrafos de dispersión, con un margen de error del 20%.

ACTIVIDAD DE

APRENDIZAJE No. 1

I. IDENTIFICACION DE LOS PRINCIPALES ESTADIGRAFOS DE DISPERSION

OBJETIVO FACILITADOR No.1

Dada una prueba escrita, al terminar la actividad de aprendizaje, usted estará en capacidad de identificar los principales estadígrafos de dispersión que verá en esta unidad, sin margen de error.

A. DEFINICION:

Los estadígrafos de dispersión son medidas que se emplean para determinar el grado de variabilidad o de dispersión de los datos con respecto a un promedio representativo de los mismos. Por ejemplo, la Desviación Media de los salarios, le indicará (en promedio) en cuantos pesos se alejan los datos de la Media de dichos salarios.

B. PRINCIPALES ESTADIGRAFOS DE DISPERSION:

> Las medidas de dispersión, por su relación con los promedios, pueden ser de mucha utilidad en el análisis.

En esta unidad usted estudiará las siguientes medidas de dispersión:

- Desviación Media
- Desviación Mediana
- Varianza
- Desviación Típica o Standard
- Puntaje Típico o Standard
- Coeficiente de Variación
- Coeficiente de Deformación o Asimetría
- Coeficiente de Apuntamiento o Curtosis.

Para facilitarle el aprendizaje de estos estadígrafos y, a su vez, hacerle notar una diferencia básica que existe entre un grupo de ellos y el resto, se ha decidido clasificarlos y analizarlos en dos bloques: El primero, correspondiente a los denominados <u>"Estadígrafos de Dispersión Absoluta"</u>, y el segundo, a los llamados <u>"Estadígrafos de Dispersión Relativa"</u>.

PRUEBA DE AVANCE No. 1

1. Defina estadígrafos de dispersión.

2. Enuncie los estadígrafos de dispersión que verá en esta unidad.

LA CORRECCION DE ESTA PRUEBA ESTARA A CARGO DEL INSTRUCTOR.

ACTIVIDAD DE
APRENDIZAJE No. 2

II. CALCULO DE ESTADIGRAFOS DE DISPERSION ABSOLUTA

OBJETIVO FACILITADOR No. 2

Dadas la clasificación y tabulación de los datos, usted estará en capacidad, al terminar la actividad de aprendizaje, de calcularle los estadígrafos de dispersión respectivos, con un margen de error del 20%.

A. DEFINICION:

Los estadígrafos de dispersión absoluta son medidas que se expresan en las mismas unidades de la variable. Por ejemplo, si la variable se expresa en pesos ($), dichos estadígrafos se expresarán igualmente en pesos, etc.

Por su mayor uso, los estadígrafos de dispersión absoluta que estudiará en esta unidad son:

- Desviación Media
- Desviación Mediana
- Varianza
- Desviación Típica o Standard.

Estos estadígrafos, exceptúando la Varianza, se expresan en las mismas unidades de la Variable.

Debemos advertirle, antes de continuar, que de todas las medidas anteriores, la única que no se expresa en las mismas unidades de la variable es la Varianza, como podrá comprobarlo más adelante.

B. CALCULO DE LA DESVIACION MEDIA:

En el cálculo de la Desviación Media, como en el de todas las demás medidas de dispersión, vamos a utilizar dos de las cuatro series que usted ha venido estudiando desde la unidad No. 2, es decir, las series No. 3 (salarios/hora) y No. 4 (figuras de porcelana dañadas/caja).

Empecemos, entonces, calculándole la Desviación Media a la serie No. 3, para lo cual haremos uso de la siguiente fórmula:

$$D_{\bar{y}} = \frac{\Sigma |Yi-\bar{Y}|\, ni}{n}$$

De acuerdo con esta fórmula, los pasos a seguir son:

1. Partir de una tabla que posea las marcas de clases (Y_i) y las frecuencias absolutas simples (ni)

2. Calcular una tercera columna con las diferencias $Yi - \bar{Y}$ en valores absolutos, es decir, considerar todas las

diferencias positivas, aunque algunas resultaren negativas.

3. Determinar una cuarta columna que sea el resultado de multiplicar las diferencias $Y_i - \bar{Y}$ por las n_i.

4. Sumar la cuarta columna

5. Dividir la suma anterior por el tamaño de la muestra (n).

O sea:

$\boxed{\bar{Y} = 88.8}$ (recuerde que este valor lo halló en la unidad No. 3, al calcular la media aritmética de la serie No. 3).

Salarios hora	Número Obreros		
Marcas de Clases	Número Obreros		
Yi	ni	Yi-$\bar{Y}$	(Yi-$\bar{Y}$)ni
68.5	2	20.3	40.6
75.5.	9	13.3	119.7
82.5	8	6.3	50.4
89.5	14	0.7	9.8
96.5	9	7.7	69.3
103.5	6	14.7	88.2
110.5	2	21.7	43.4
	50		421.4

$$D\bar{Y} = \frac{\Sigma (y_i - \bar{Y})\, n_i}{n}$$

$$D\bar{y} = \frac{421.4}{50}$$

$$\boxed{D\bar{y} = 8.43}$$

El valor que acabamos de hallar nos indica que "los salarios se alejan de su media aritmética en $8.43 (en promedio)".

Si observa detenidamente la fórmula de la Desviación Mediana, notará su similitud con la de la Media Aritmética, puesto que ambas nos dicen que debemos sumar (Σ) y luego dividir por el tamaño de la muestra (n). Por tal razón se concluye que:

"La Desviación media es la media de las desviaciones (o diferencias $Y_i - \bar{Y}$) con respecto al promedio aritmético, expresadas en valores absolutos".

La razón de expresar las diferencias $Y_i - \bar{Y}$ en valores absolutos, se debe a la propiedad de la Media Aritmética que usted estudió en la anterior unidad, según la cual

$\Sigma (Y_i - \bar{Y})\, n_i = 0$

Las diferencias $Y_i - \bar{Y}$ están en valores absolutos, porque $\Sigma (Y_i - \bar{Y})\, n_i = 0$

En efecto, note que si prescinde de los valores absolutos y considera los productos $(Y_i - \bar{Y})\, n_i$ con signos - y +, la suma de los productos negativos se anulará con la correspondiente a los positivos.

En datos no agrupados, la fórmula para calcular la Desviación Media es: $D_{\bar{X}} = \frac{\Sigma [X_i - \bar{X}]}{n}$

EJERCICIO:

Calcule e interprete la Desviación Media de la Serie No. 4.

C. CALCULO DE LA DESVIACION MEDIANA:

El cálculo de esta medida le será más fácil, si asimiló bien el del estadígrafo anterior, puesto que como comprobará, su fórmula sugiere un procedimiento similar. En efecto, para hallar la Desviación Mediana usted deberá emplear la siguiente expresión para datos agrupados:

$$D_{Me} = \frac{\Sigma [Y_i - Me]\, n_i}{n}$$

Observe que la única diferencia con la fórmula de la Desviación Media, consiste en que las diferencias no son $Y_i - \bar{Y}$, sino $Y_i - Me$. Por lo tanto, los pasos a seguir en el cálculo de esta medida serán los mismos que siguió para hallar a la Desviación Media.

Ejemplo:

Hallar la Desviación Mediana de la serie Salarios/hora:

Solución: Me = 89 (este valor lo halló en la unidad anterior al calcular la Mediana).

Marcas de clase	Número Obreros		
Yi	n_i	Yi -Me	(Yi-Me) ni
68.5	2	20.5	41.0
75.5	9	13.5	121.5
82.5	8	6.5	52.0
89.5	14	0.5	7.0
96.5	9	7.5	67.5
103.5	6	14.5	87.0
110.5	2	21.5	43.0
	50		419.0 *

$$D_{Me} = \frac{\Sigma |Yi - Me| ni}{n}$$

$$D_{Me} = \frac{419}{50}$$

$$\boxed{D_{Me} = 8.38}$$

Este valor indica que "los salarios difieren (en promedio) en \$8.38 de su Mediana"

Para datos no agrupados la Desviación Mediana es:

$$D_{Me} = \frac{\Sigma |X_i - Me}{n}$$

EJERCICIO:

Calcule e interprete la Desviación Mediana de la serie No.4

D. CALCULO DE LA VARIANZA:

La varianza y la Desviación Típica o Standard (que estudiará inmediatamente después), son los estadígrafos de dispersión que más se usan, razón por la cual se le recomienda especial atención y dedicación en su estudio.

En esta parte, usted iniciará con el aprendizaje del cálculo de la varianza, a quien los autores de textos estadísticos simbolizan como S^2_y, o, S^2_x, o, $V[Y]$, o $V[X]$, o, $M[Z^2_i]$.

La varianza y la desviación típica son dos medidas útiles y están ligadas entre sí.

Las fórmulas para calcular la varianza son:

a) $$S^2_x = \frac{\Sigma(X_i-\bar{X})^2}{n} = \frac{\Sigma Z^2}{n}$$, si

los datos no están agrupados

* b) $$S^2_y = \frac{\Sigma(Y_i-\bar{Y})^2 n_i}{n} = \frac{\Sigma Z_i^2 n_i}{n}$$, si

los datos están agrupados.

EJEMPLO No. 1:

Supongamos que las ventas diarias de un almacén durante una semana son las siguientes (en miles de $).

Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
60	80	88	98	106	120

Se pide calcular la varianza de las ventas.

SOLUCION:

Como la serie anterior es poco numerosa y sus datos no están agrupados, usted debe utilizar la fórmula:

$$S^2_x = \frac{\Sigma(X_i - \bar{X})^2}{n}$$

Se le recomienda, por lo tanto, seguir los siguientes pasos:

1. Calcular a $\bar{X}$ (media aritmética)

2. Obtener las diferencias $X_i-\bar{X}$, es decir, restarle a cada uno de los valores de la serie el promedio aritmético.

3. Elevar al cuadrado las diferencias $X_i - \bar{X}$

4. Sumar las $(X_i - \bar{X})^2$

5. Dividir la suma de las $(Xi-\bar{X})^2$ por el tamaño de la muestra (esto es, por $n = 6$).

O sea:

$$\bar{X} = \frac{60 + 80 + 88 + 98 + 106 + 120}{6} = \frac{552}{6} = 92$$

Entonces:

	Z_i	Z_i^2
Xi	$Xi-\bar{X}$	$(Xi-\bar{X})^2$
60	-32	1.024
80	-12	144
88	- 4	16
98	+ 6	36
106	+14	196
120	+28	784
	0	2.200

$$S^2_X = \frac{2.200}{6} = 366.66$$

Observe que $\Sigma\, Zi=0$, razón por la cual se trabaja con $\Sigma\, Zi^2$.

EJEMPLO No. 2

Calcular la varianza de la serie No. 3 (salarios/hora).

SOLUCION:

En este caso los datos están agrupados, por lo que se debe usar la fórmula:

$$S^2_y = \frac{\Sigma (Y_i - \bar{Y})^2 n_i}{n} \quad *$$

En consecuencia:

$\bar{Y}$ = 88.8 (esta media se halló para los salarios/hora en la unidad anterior sobre Medidas de Tendencia Central).

		Z_i	Z_i^2	$Z_i^2 n_i$
Y_i	n_i	$Y_i-\bar{Y}$	$(Y_i-\bar{Y})^2$	$(Y_i-\bar{Y})^2 n_i$
68.5	2	-20.3	412.09	824.18
75.5	9	-13.3	176.89	1.592.01
82.5	8	- 6.3	39.69	317.52
89.5	14	+ 0.7	0.49	6.86
96.5	9	+ 7.7	59.29	533.61
103.5	6	+14.7	216.09	1.296.54
110.5	2	+21.7	470.89	941.78
	50			5.512.50

$$S^2_y = \frac{5.512.5}{50} = 110.25$$

Note que la varianza de esta serie se ha calculado también siguiendo los pasos descritos antes para datos no agrupados, pero se agregó un sexto paso que consiste en multiplicar a las $(Y_i-\bar{Y})^2$ por n_i.

Si analiza las fórmulas anteriores de la varianza, se dará cuenta porqué se la define como "la media de las desviaciones al cuadrado con respecto al promedio aritmético".

EJERCICIO:

Calcule la varianza de la serie No. 4 (figuras de percelana dañadas).

Compare sus respuestas con las del Instructor.

Otros métodos de cálculo de la varianza son los siguientes:

1. $S^2_x = \frac{\Sigma X_i^2}{n} - (\bar{X})^2,$

para datos no agrupados.

2. $S^2_y = \frac{\Sigma Y_i^2 n_i}{n} - (\bar{Y})^2,$

para datos agrupados.

Existen otros métodos para calcular la varianza, ¿Cómo serán?

Además de las anteriores, existen otras formas de calcular la varianza, conocidas con el nombre de "Métodos abreviados".

Si está interesado en conocer y aplicar los anteriores métodos, le recomendamos consultar la Unidad No.8 de la Estadística Elemental de Horacio D'Ottone, o las unidades de Estadística del SENA (ediciones anteriores).

La varianza, como la media aritmética, posee algunas propiedades que resultan útiles en el desarrollo de ciertos ejercicios y problemas. Entre tales propiedades cabe destacar las siguientes:

a. La varianza de una constante es igual a cero

$$V[K] = 0$$

EJEMPLO:

Suponga que en la primera evaluación de estadística todos los 20 alumnos sacaron la misma nota, verbigracia, 6 puntos (sobre 10). Se pide entonces calcular la varianza de las notas.

SOLUCION:

La nota promedio es 6, es decir:

$\bar{X} = 6$, porque todos obtuvieron igual nota.

Si aplica ahora la fórmula de la varianza para datos no agrupados, tendrá que:

$$S^2_X = \frac{\sum (X_i - \bar{X})^2}{n} = \frac{(6-6)^2+(6-6)^2+(6-6)^2+\ldots+(6-6)^2}{20} = \frac{0}{20} = 0$$

b. La varianza de la suma (o resta) de una variable más (o menos) una constante, es igual a la varianza de la variable:

$$V[X+K] = V[X] + V[K]$$

Como $V[K] = 0$, entonces:

$$\boxed{V[X+K] = V[X]}$$

La constante que le sumo o resto a cada uno de los datos no altera la varianza.

Tambien:

$$\boxed{V[X-K] = V[X]},$$

por la misma razón.

EJEMPLO:

Ochenta empleados de una compañía tienen un salario promedio de $14 500. y la varianza de 1.200.

Cuando la compañía divulgó dicho promedio, el sindicato de la empresa protestó porque en él se había incluido bonificación de $2.000. que se pagó únicamente en ese mes, por haber cumplido la empresa 20 años de fundada. Se pide rectificar las anteriores medidas.

SOLUCION:

$\bar{X} = 14.500$

$S^2_x = 1.200$

Como el promedio incluyó incorrectamente los $2.000. de bonificación, su rectificación se hace mediante la propiedad siguiente:

$M\ [X - K] = \bar{X} - K = 14.500 - 2.000 = 12.500.$

La inclusión de la bonificación exige igualmente la rectificación de la varianza, así:

$V\ [X-K] = V\ [X] = 1.200.$

O sea que el promedio se había afectado con la inclusión de la bonificación, pero la varianza no.

Estas propiedades son parecidas a las de la Media Aritmética.

c. La varianza de una constante por una variable, es igual al producto de la constante al cuadrado por la varianza de la variable: $\boxed{V\ [K\ X] = K^2 .\ V\ [X]}$

EJEMPLO:

La compañía acepta el reclamo del sindicato y ordena además, en compensación, un reajuste del 20% en cada uno de los salarios. Se pide calcular la nueva varianza.

SOLUCION:

$S^2_x = 1.200$

El reajuste del 20% da origen a una constante igual a 1.20, que se deriva del hecho de que cada peso ganado por cada obrero se convierte, con el aumento, en \$1.20. O sea que: $K = 1.20$

Por lo tanto, aplicando la propiedad anterior queda que:

$V[K.X] = K^2 . V[X] = (1.20)^2 \times 1.200$

$V[K.X] = 1.44 \times 1.200 = 1.728$

EJERCICIO:

Suponga que en la empresa donde trabajan los 50 obreros de la serie No. 3 (salarios/hora), le proponen a la comisión negociadora del Sindicato las siguientes soluciones:

a. Aumento general de \$22 por hora para todos los obreros.

b. Aumento general del 25%.

Calcule la nueva varianza en cada caso y determine cuál de las dos propuestas es más favorable para los obreros.

E. CALCULO DE LA DESVIACION TIPICA O STANDARD:

La desviación típica o standard se define como "la raíz positiva de la varianza". Esto es: $S = + \sqrt{S^2}$

Para calcular la desviación típica, necesito la varianza.

EJEMPLO No.1:

Calcular la desviación típica de la serie sobre las ventas diarias durante una semana que se analizó en el cálculo de la varianza.

SOLUCION:

$S^2_x = 366.66$

$Sx = + \sqrt{S^2_x} = + \sqrt{366.66} = + 19.15$

El resultado nos dice que la desviación típica de las ventas diarias durante la semana es de $19.15.

EJEMPLO No. 2:

Calcular la desviación típica de los salarios/hora (serie No.3).

SOLUCION:

S^2_y = 110.25 (hallada en el punto anterior)

$Sy = + \sqrt{S^2y} = + \sqrt{110.25} = 10.5$

Quiere decir tal resultado que la desviación típica de los salarios/hora es de $10.5.

EJERCICIO:

Calcule e interprete la desviación típica de las figuras de porcelana dañadas (serie No. 4).

Compare sus respuestas con las del Instructor.

Recuerda la forma de la Distribución Simétrica o Normal, vista en la unidad sobre Medidas de Tendencia Central?.

La Distribución Normal tiene forma de campana.

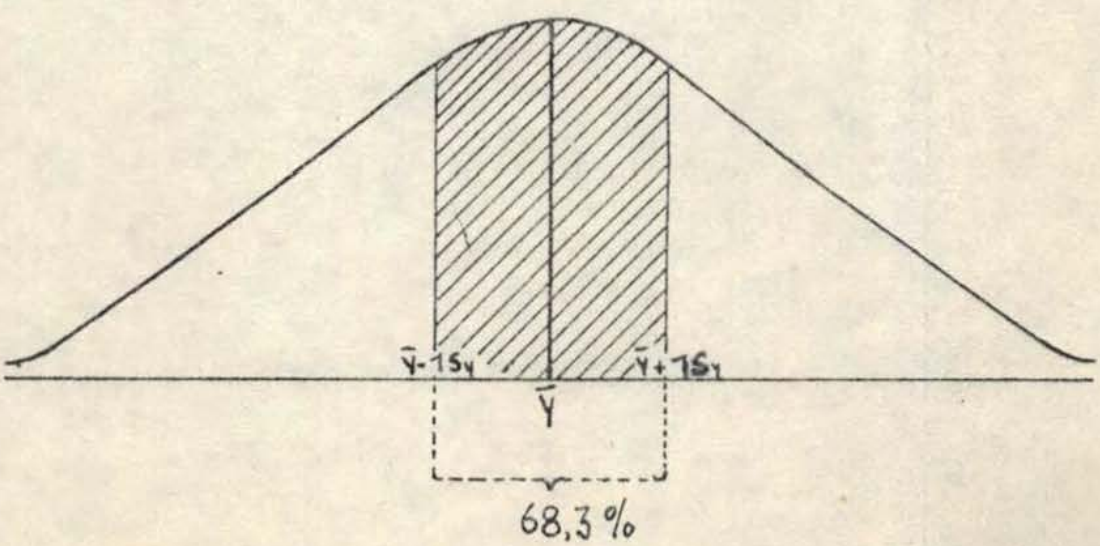
¿qué tendrá que ver la desviación típica con ella.?

Si no la recuerda bien, le

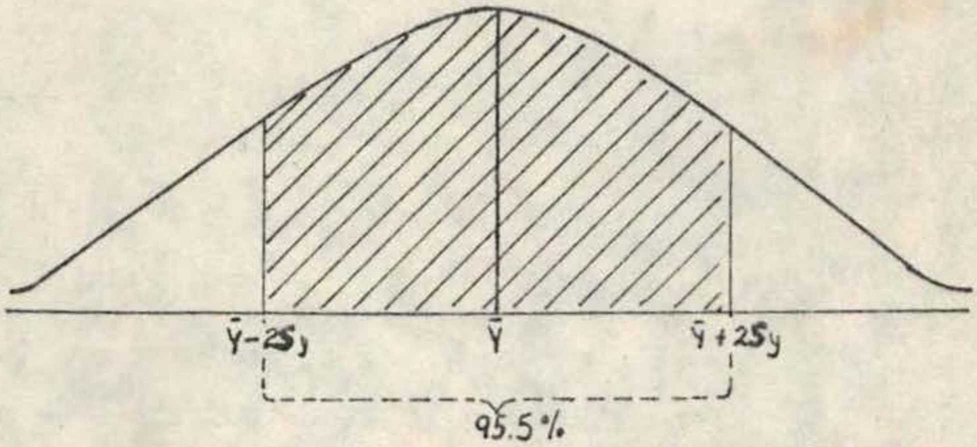
sugerimos echarle nuevamente una mirada detenida, pues la Desviación Típica o Standard tiene en relación con dicha distribución ciertas propiedades de suma importancia, que son igualmente aplicables a cualquier otra distribución no simétrica.

Estas propiedades se derivan de una desigualdad llamada de TChebycheff, que debe su nombre a un investigador ruso y que verá en detalle más adelante en la Teoría de muestreo. Algunas de ellas son las siguientes:

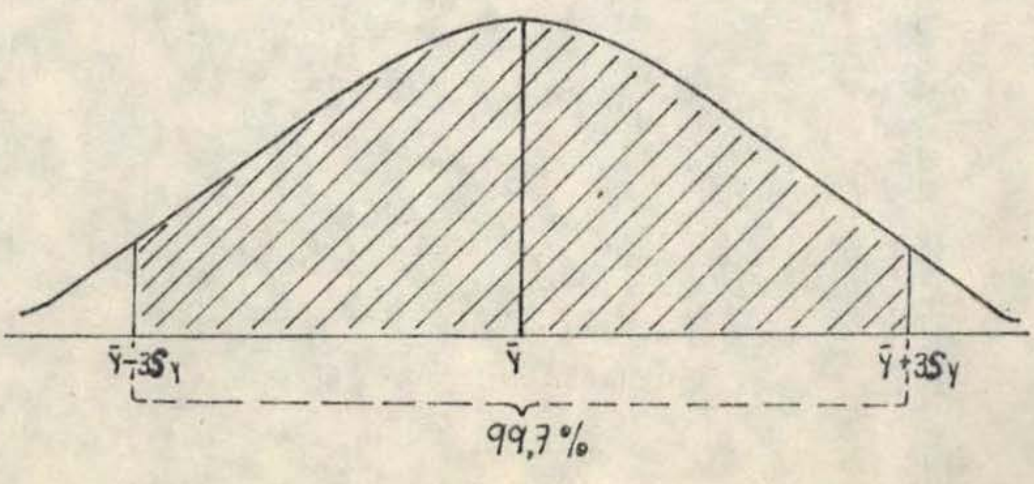
1. En un intervalo de la Curva Normal, cuyos extremos sean la media aritmética menos una vez la desviación típica, el inferior, y la media aritmética más una vez la desviación típica, el superior, se encuentra el 68.3 % de los datos de la serie. Esto es:



2. En un intervalo, cuyos extremos sean la media aritmética menos dos veces la desviación típica (el límite inferior), y la media más dos veces la desviación típica (el límite superior), se encuentra el 95.5% de las observaciones. Es decir:



3. En un intervalo, cuyos extremos sean la media aritmética menos tres veces la desviación típica (el inferior) y la media más tres veces la desviación típica (el superior), se encuentra el 99.7% de las observaciones. O sea:



Estas propiedades son muy usadas en el área educativa, a nivel de entidades como ICFES e ICETEX, por ejemplo.

Desarrollaremos algunos casos para que aprenda a utilizarlas:

EJEMPLO No. 1:

Entre qué límites de confianza se encuentra el 68.3% de las observaciones correspondientes a la muestra de los salarios/hora (Serie No. 3)?

SOLUCION:

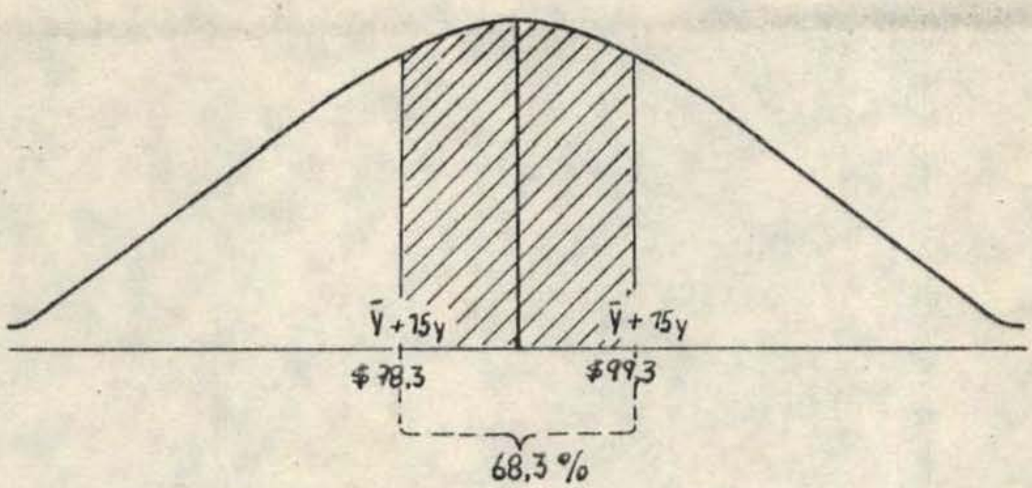
Sabemos que para la serie salarios/hora $\bar{Y} = 88.8$ y $Sy=10.5$. Se ha dicho también que el intervalo que recoge el 68.3% de las observaciones está limitado por los extremos $\bar{Y} - 1S_y$ y $\bar{Y} + 1Sy$, por lo tanto:

$\bar{Y} - 1Sy = 88.8 - 1 \times 10.5 = 88.8 - 10.5 = 78.3$

$\bar{Y} + 1Sy = 88.8 + 1 \times 10.5 = 88.8 + 10.5 = 99.3$

O sea que el 68.3% de los salarios/hora se encuentra entre los valores \$78.30 y \$99.30, que son los límites de confianza.

Representemos gráficamente esta situación así:



EJEMPLO No. 2:

Cuántos salarios/hora se espera hallar con el grado de confianza del 68.3% en el intervalo correspondiente (esto es, entre \$78.3 y \$99.3)?.

SOLUCION:

El número de datos que se espera hallar en dicho intervalo debe ser igual al 68.3% del total de salarios/hora, es decir el 68.3% de las 50 observaciones de la muestra, esto es:

50 x 68.3% = 50 x 0.683 = 34.15

Es decir, se espera encontrar 34 datos en el intervalo comprendido entre los salarios \$78.30 y \$99.30.

Si contamos en la serie sin agrupar (unidad No. 2) los salarios/hora comprendidos entre esos dos valores, el resultado real será 31, que difiere de lo esperado en 3 observaciones, debido a que la serie no es simétrica. De todas formas, el resultado es satisfactorio, lo cual es indicativo de la confiabilidad que arrojan las propiedades descritas.

EJERCICIO:

Suponga que con un grado de confiabilidad del 95.5% se desea conocer en la serie de salarios las siguientes características de la muestra:

a. Límites de confiabilidad
b. Número de observaciones esperadas
c. Número real de datos entre los límites de confianza del intervalo correspondiente.

Cuáles son sus respuestas?.

COMPARE SUS RESULTADOS CON LOS DEL INSTRUCTOR.

PRUEBA DE AVANCE No. 2

Los 100 empleados de una compañía se han clasificado de acuerdo con los impuestos pagados (retención en la fuente), como lo muestra la siguiente tabla:

RETENCION (Cientos $	No.EMPELADOS ni
0 - 20	30
20 - 40	25
40 - 60	15
60 - 80	13
80 - 100	12
100 - 120	5

1) Se pide calcular e interpretar

   a) Desviación Media
   b) Desviación Mediana
   c) Varianza
   d) Desviación Típica.

2) Si se reajusta el impuesto en un 10% con base en la clasificación anterior, cuál será la nueva desviación típica?.

LA CORRECCION DE ESTA PRUEBA ESTARA A CARGO DEL INSTRUCTOR.

ACTIVIDAD DE

APRENDIZAJE No. 3

II. CALCULO DE ESTADIGRAFOS DE DISPERSION RELATIVA:

OBJETIVO FACILITADOR No.3

Dada una clasificación y tabulación de los datos, al terminar la actividad de aprendizaje, usted estará en capacidad de calcular los estadígrafos de dispersión relativa, con un margen de error del 20%.

A. DEFINICION:

Las medidas de dispersión relativa son valores que se expresan en porcentajes o en fracciones de la unidad. Entre los de mayor uso veremos:

- El puntaje típico
- Coeficiente de variación
- Coeficiente de deformación o Asimetría

- Coeficiente de apuntamiento o Curtosis.

B. PUNTAJE TIPICO:

Estas medidas no se expresan, como las anteriores, en pesos, libras, hijos, etc.

Es el coeficiente que expresa la desviación de cualquier dato de la serie con respecto a la media aritmética, en unidades de desviación típica. Se simboliza por Z y la fórmula para su cálculo es:

$$Z = \frac{X_i - \bar{X}}{Sx}$$

Esta medida se emplea para comparar dos o más datos individuales, aunque pertenezca a distribuciones diferentes. Igualmente este estadígrafo le será de mucha utilidad cuando tenga que estudiar la unidad sobre Relación y Correlación Lineal, o las áreas de una Distribución Normal en la Teoría de Muestreo y Proyecciones.

EJEMPLO:

Una empresa fabrica bombillas eléctricas de dos clases, A y B. Con base en muestras de la producción se sabe que las distribuciones de la duración en horas de esas bombillas son tales que tienen las siguientes medias y varianzas:

TIPO	MEDIA	VARIANZA
A	800 horas	7.800
B	650 horas	5.400

Si se extrajo una bombilla de cada tipo y su duración fue de 700 y 630 horas, respectivamente, cuál tipo de bombilla tiene mejor posición relativa?.

SOLUCION:

Este problema se resuelve con el puntaje típico o estandarizado, así:

Para las bombillas tipo A:

$\bar{X} = 800$ horas

$S^2_x = 7.800 \qquad S_x = +\sqrt{S^2_x} = +\sqrt{7.800} = 88.32$

$X_1 = 700$ horas

Con el puntaje típico puede determinar qué bombilla se comportó relativamente mejor

Por lo tanto:

$$Z = \frac{X_1 - \bar{X}}{S_x} = \frac{700-800}{88.32} =$$

$$= \frac{-100}{88.32}$$

$$\boxed{Z = -1.13}$$

Para las bombillas tipo B:

$\bar{X} = 650$ horas

$S^2_x = 5.400 \qquad S_x = +\sqrt{S^2_x} = +\sqrt{5.400} = 73.48$

$X_1 = 630$ horas

En consecuencia:

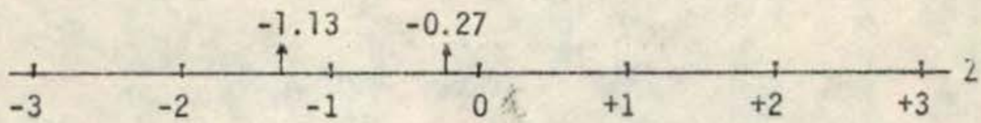
$$Z = \frac{X_1 - \bar{X}}{Sx} = \frac{630-650}{73.48} = \frac{-20}{73.48} = -0.27$$

CONCLUSION:

Decimos entonces que la bombilla del tipo B tiene una mejor posición relativa que la de tipo A, puesto que presenta un desvío respecto a la media de apenas - 0.27 veces la desviación Standard, mientras que en la de tipo A es de -1.13.

En otras palabras, por estar - 0.27 más cerca de cero (0) que -1.13, la situación de la bombilla de tipo B es mejor que la de tipo A, puesto que es sabido que entre dos números negativos es mayor el que se encuentre más próximo al cero (o); pero la situación se hubiera invertido si los dos resultados fuesen positivos, esto es, sería mejor la posición de la bombilla de tipo A que la de B.

A propósito, conviene que recuerde que el puntaje típico toma valores de -3 a 0 y de 0 a +3, lo cual se representa gráficamente así:



No olvide tampoco que entre mayor sea el valor de este puntaje, mejor será la posición relativa del elemento que lo obtenga.

EJERCICIO:

Suponga que en el primer parcial de estadística el grupo obtuvo sobre un puntaje de 10, una media general de 4,4 y una desviación típica de 0.34; en tanto que en el primer parcial de economía, obtuvo una media de 5.3 y desviación típica de 0.57.

Si se me presentan dos puntajes positivos, diré que tiene mejor posición relativamente el de mayor puntaje

Si un estudiante A obtuvo 5.5 en estadística y otro estudiante B obtuvo 6.4 en economía, ¿Quién se desempeñó relativamente mejor?.

COMPARE SUS RESPUESTAS CON LAS DEL INSTRUCTOR

C. COEFICIENTE DE VARIACION:

Es el cociente que compara a la desviación típica con la media aritmética. Se expresa en porcentaje y lo simbolizaremos por CV, así:

$$CV = \frac{Sx}{\bar{X}} \times 100\%, \text{ o también:}$$

$$CV = \frac{Sy}{\bar{Y}} \times 100\%$$

Este coeficiente se emplea cuando se desea comparar dos o más distribuciones, con el fin de determinar cual de ellas tiene mayor o menor variabilidad relativa. Su uso se hace necesario cuando dichas distribuciones están dadas en unidades diferentes (por ejemplo, hijos y salarios) y, por lo tanto, su comparación no se puede hacer con los otros estadígrafos antes estudiados.

EJEMPLO:

Con este coeficiente puedo comparar la variabilidad de dos series diferentes.

Supongamos que un grupo de profesionales, que trabaja en un sector de la actividad industrial de un país A, tiene un salario promedio de $36.800 con una varianza de los salarios de

144.000, mientras que otro grupo de empleados que trabajan en un país B, en una actividad similar, tiene un salario promedio de 8.570 bolívares y la desviación típica de los salarios es de 800 bolívares. Se quiere determinar cual grupo de salarios presenta una menor variabilidad.

SOLUCION:

Para el grupo de profesionales del país A, se tiene que:

$\bar{X}$ = \$36.800

$S^2_x = 144.000 \qquad Sx \sqrt{S^2_x} = \sqrt{144.000} = 379.47$

Por lo tanto:

$$CV = \frac{Sx}{\bar{X}} \times 100\% = \frac{379.47}{36.800} \times 100\% = 1.03\ \%.$$

Para el grupo de empleados del país B, tenemos:

$\bar{X}$ = 8.750 bolívares

Sx = 800 bolívares

En consecuencia:

$$CV = \frac{Sx}{\bar{X}} \times 100\% = \frac{800}{8.750} \times 100\ \% = 9.14\%$$

CONCLUSION:

Los salarios para el grupo de empleados profesionales del país A presenta menor variabilidad que los salarios de los empleados del país B.

Si el CV<30% la serie es homogénea.
Si el CV>30%, la muestra es heterogénea.

Se dice que si una serie tiene un coeficiente de variación menor que el 30%, es homogénea o poco dispersa respecto a su promedio aritmético; en tanto que una muestra cuyo coeficiente de variación sea mayor que el 30%, se dice que es heterogénea o dispersa en relación con su promedio.

Se afirma, además, que una serie con un coeficiente de variación superior al 50% es una muestra muy dispersa y que, en consecuencia, los estadígrafos que recogerían a todos sus datos, como la media aritmética y la varianza, no serían representativos de los datos que la constituyen.

De lo anterior se deriva que los salarios de los dos grupos de empleados analizados en el ejemplo que acabamos de ver, constituyen series homogéneas o concentradas alrededor de sus respectivos promedios aritméticos, puesto que sus coeficientes de variación son menores que 30% y, además bastante pequeños.

EJERCICIO:

En cierta región, la distribución de predios por extensión tiene una media de 35.4 hectáreas y una desviación típica de 19.33 hectáreas, mientras que la distribución por canon de arrendamiento tiene una media de $30.750 y una desviación típica de $4.590. ¿Cuál distribución tiene mayor variabilidad? y ¿Qué tipo de serie es cada una, de acuerdo con el valor de su coeficiente de variación?.

Debo hallarle el CV a cada distribución para saber cuál tiene mayor variabilidad.

Antes de pasar al estudio del próximo estadígrafo de dispersión, conviene que recuerde lo siguiente:

El cálculo de la variación relativa, utilizando el coeficiente de variación, es criticado por el inconveniente que presenta cuando dos distribuciones con medias aritméticas diferentes pero iguales desviaciones típicas, es decir, con la misma variabilidad, arrojan diferentes resultados.

Consideremos, por ejemplo, dos distribuciones cuyas medias aritméticas son 24.5 y 30, y cuyas desviaciones típicas o estandars son iguales a 2 (indicándonos este hecho que tienen el mismo grado de variación absoluta).

Al calcular, entonces, sus coeficientes de variación nos encontramos, sin embargo, con que éstos son diferentes, es decir, que las variaciones relativas de las dos distribuciones son distintas, a pesar de ser iguales sus variaciones absolutas, lo cual es una contradicción.

Veamos:

Para la primera distribución:

Ciertamente, es una contradicción que tales distribuciones tengan iguales desviaciones típicas, pero diferentes coeficientes de variación.

$\bar{X}_1 = 24.5$

$S_X = 2$

Entonces:

$$CV_1 = \frac{S_X}{\bar{X}} \times 100\% = \frac{2}{24.5} \times 100\%$$

$$\boxed{CV_1 = 8.16\%}$$

Para la segunda distribución:

$\bar{X}_2 = 30$

$S_X = 2$

Por lo tanto:

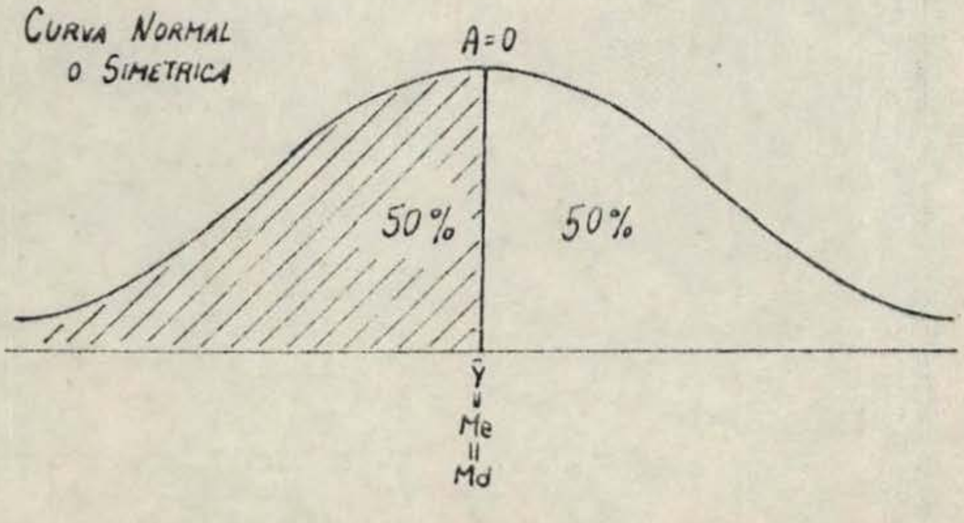
$$CV_2 = \frac{Sx}{\bar{X}_2} \times 100\% = \frac{2}{30} \times 100\%$$

$$\boxed{CV_2 = 6.66\%}$$

O sea que $CV_1 \neq CV_2$

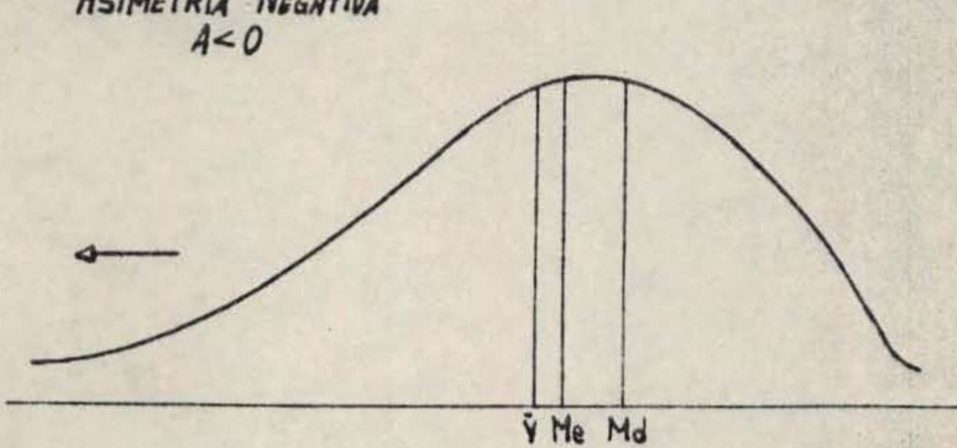
D. COEFICIENTE DE DEFORMACION O ASIMETRIA:

Este coeficiente trata de explicar el grado de deformación de la distribución, comparándola con una distribución teórica normal. En esta última, la asimetría es igual a cero, porque sus datos están distribuidos en la misma proporción hacia la izquierda y hacia la derecha de su media aritmética, que resulta también igual a la moda y a la mediana. Esto es:



Esta curva será nuestro punto de partida para analizar la asimetría que presenten las demás distribuciones que tengamos que estudiar y cuyas formas, generalemtne, no se asemejan a la de una distribución normal.

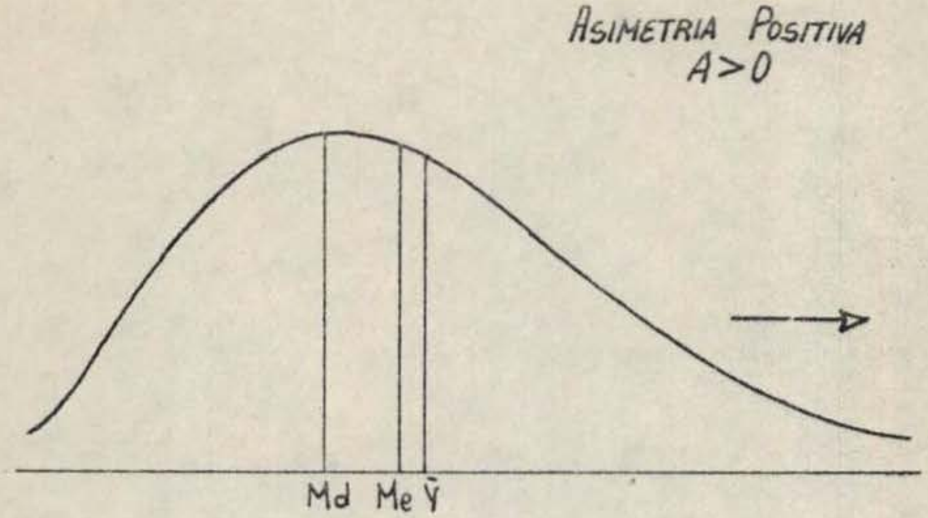
Hay distribuciones que son asimétricas hacia la izquierda, lo cual se reconoce por la dirección de su cola, así:



En estas distribuciones la asimetría es menor que cero, por lo cual se afirma que tienen asimetría negativa. Note, igualmente, que en ellas la media aritmética es menor que la mediana y ésta, a su vez, es menor que la moda, o sea:

$$\boxed{\bar{Y} < Me < Md.}$$

Otras distribuciones son asimétricas hacia la derecha, es decir, tienen asimetría mayor que cero o positiva y su representación gráfica es la siguiente:



En tales distribuciones la media aritmética es mayor que la mediana y ésta mayor que la moda, esto es:

$$\boxed{\bar{Y} > Me > Md}$$

En la práctica, algunas entidades que elaboran información estadística, calculan la asimetría negativa (cuando $\bar{Y} < Me < Md$) con la siguiente fórmula:

$$A_1 = \frac{\bar{Y} - Md}{Sy}$$

Esas mismas entidades utilizan otra fórmula para distribuciones positivas (cuando $\bar{Y} > Me > Md$). Tal fórmula es la siguiente:

$$A_2 = \frac{3\ (\bar{Y} - Me)}{Sy}$$

Existen, sin embargo, otras fórmulas para calcular la asimetría. Consulténse, por ejemplo, los siguientes textos:

- Estadística Comercial de Ciro Martínez (Editorial Norma), Unidad No. 8, sección 8.3.

- Estadística Elemental de Horacio D'Ottone (Editorial Cienes), capítulo No. 4, sección 4.8.

Veamos ahora algunos casos de análisis de asimetría.

EJEMPLO:

Calcular la asimetría de la distribución de los salarios/hora (serie No.3).

SOLUCION:

$\bar{Y} = 88.8$

$Me = 89$

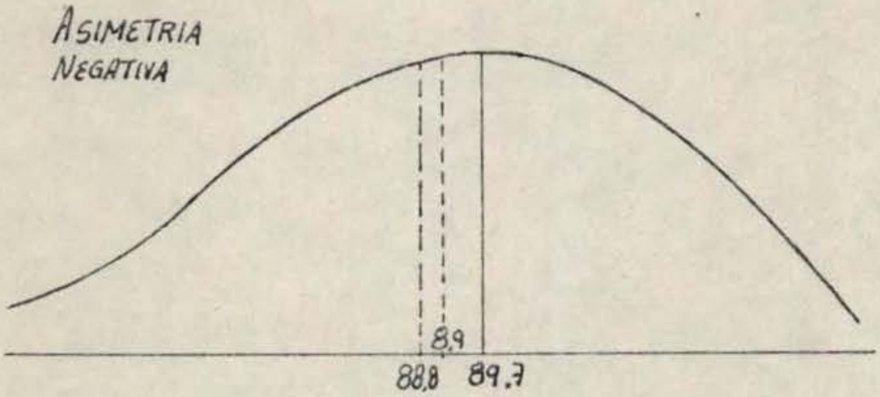
$Md. = 89.7$

$Sy = 10.5$

Como $\bar{Y} < Me < Md$, se supone que la distribución de los salarios tiene una asimetría negativa $(A < 0)$. Por lo tanto, se calcula con la primera fórmula, así:

$$A = \frac{\bar{Y} - Md}{Sy} = \frac{88.8 - 89.7}{10.5} = -0.0857$$

El resultado anterior nos indica que la distribución de los salarios/hora tiene una ligera asimetría hacia la izquierda, lo cual implica que los salarios están concentrados hacia los de mayor valor. Su representación gráfica, en consecuencia, podría ser la siguiente:



EJERCICIO:

Calcule e interprete la asimetría de la distribución de las figuras de porcelana desperfectas en las 30 cajas de madera (Serie No. 4). Elabore la gráfica correspondiente, de acuerdo con la asimetría.

COMPARE SUS RESPUESTAS CON LAS DEL INSTRUCTOR.

E. COEFICIENTE DE APUNTAMIENTO O CURTOSIS:

Este coeficiente trata de cuantificar el grado de homogeneidad o heterogeneidad de los datos respecto de sus valores centrales, comparando también la distribución estudiada con la distribución normal.

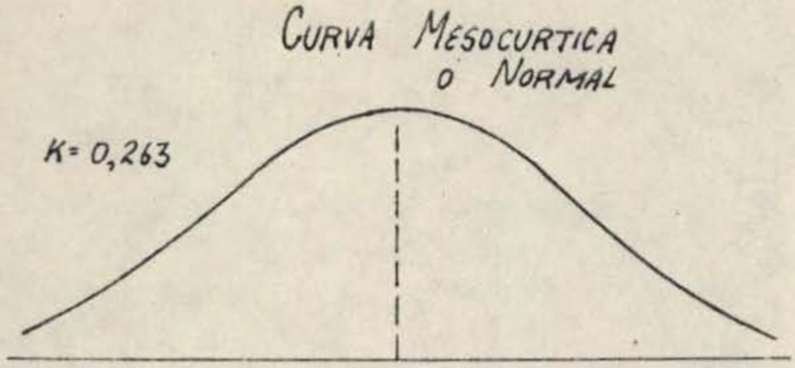
En otras palabras, mientras la asimetría le dice hacia qué lado (izquierdo o derecho) se concentran los datos, el coeficiente de apuntamiento o curtosis le informa cuán mucha o poca es la concentración o la altura de la distribución, sin importar hacia qué lado se da

El cálculo de la Curtosis se hace con unas cuantas fórmulas, de las cuales se ha seleccionado, para los fines de esta unidad, la siguiente:

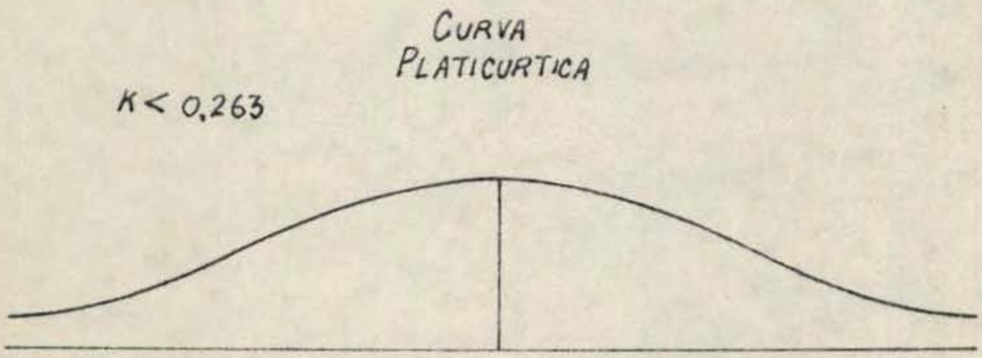
$$K = \frac{{}^{1}/_{2}\,(Q_3 - Q_1)}{D_9 - D_1}$$

Si usted está interesado, sin embargo, en conocer otras, le recomendamos consultar el texto de Ciro Martínez (obra citada), unidad 8, sección 8.4.

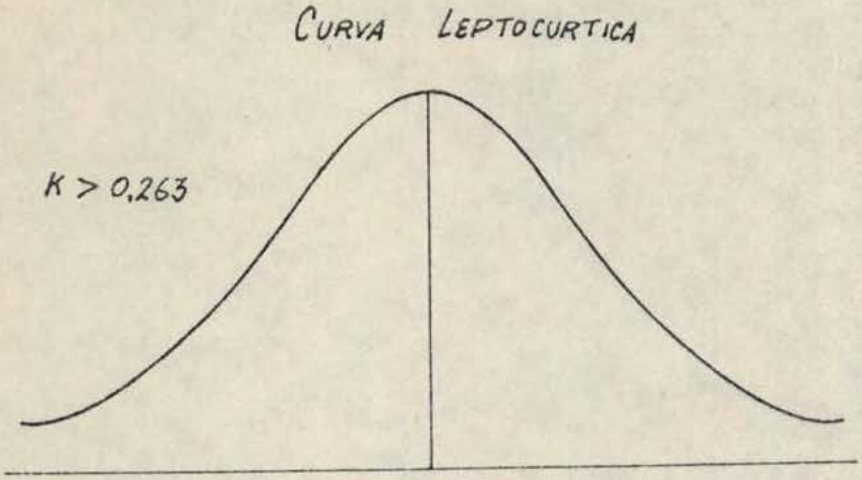
De acuerdo con la fórmula escogida, una distribución normal resulta con una curtosis igual a 0.263 y se le denomina mesocúrtica. La siguiente es su representación gráfica:



Cuando una distribución tiene una curtosis menor que o.263, se le denomina Platicúrtica o Aplanada (muy dispersa) y se le puede representar así:



Si por el contrario, una distribución tiene un coeficiente de apuntamiento mayor que 0.263, se le llama Leptocúrtica o Concentrada. Su gráfica es la siguiente:



EJEMPLO:

Calcular e interpretar el coeficiente de apuntamiento de la distribución de los salarios/hora.

SOLUCION:

$Q_3 = 96.5$

$Q_1 = 80.31$

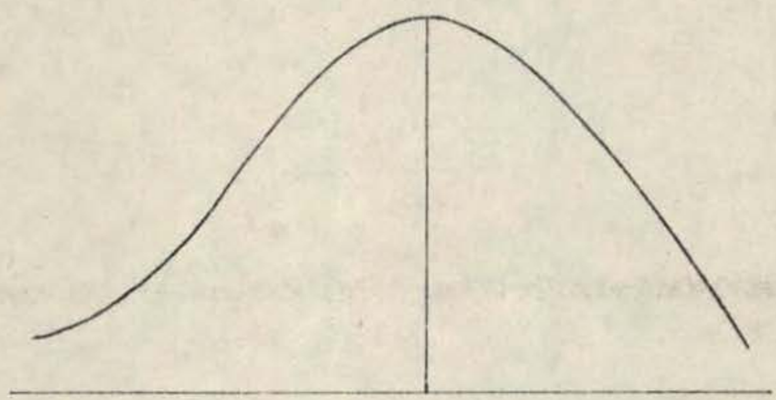
$D_9 = 103.5$

$D_1 = 74.33$

$$K = \frac{1/2(Q_3 - Q_1)}{D_9 - D_1} = \frac{0.5\ (96.50 - 80.31)}{103.50 - 74.33} = \frac{0.5 \times 16.19}{29.17} =$$

$$K = \frac{8.095}{29.17} = 0.2775$$

En este caso la curtosis es mayor que 0.263, por lo tanto la distribución es más apuntada o concentrada que una distribución normal, es decir, es ligeramente leptocúrtica. Su gráfica, teniendo en cuenta también la dirección de su cola vista en el cálculo de la asimetría, podría ser la siguiente:



EJERCICIO:

Calcule e interprete el coeficiente de apuntamiento de la distribución de las figuras de porcelana desperfectas en las 30 cajas de madera (serie No. 4)

COMPARE SUS RESPUESTAS CON LAS DEL INSTRUCTOR.

PRUEBA DE AVANCE No. 3

Partiendo de la distribución cuya clasificación y tabulación aparece en el test No. 2 de la presente unidad, calcule e interprete:

1) Coeficiente de Variación.

2) Coeficiente de Deformación o Asimetría

3) Coeficiente de Apuntamiento o Curtosis.

LA CORRECCION DE ESTA PRUEBA ESTAR A CARGO DEL INSTRUCTOR.

PRUEBA FINAL

Las edades de los estudiantes de complementación del centro de Administración SENA, se han agrupado como sigue: partiendo de una muestra de 100 alumnos:

EDAD (años)	No. ESTUDIANTES n_i
18.1 - 22	25
22.1 - 26	35
26.1 - 30	20
30.1 - 34	15
34.1 - 38	5
	100

Se pide calcular e interpretar:

1) Desviación Media
2) Desviación Mediana
3) Desviación Típica
4) Coeficiente de Variación
5) Coeficiente de Deformación o Asimetría.

LA CORRECCION DE ESTA PRUEBA ESTARA A CARGO DEL INSTRUCTOR.

BIBLIOGRAFIA

MARTINEZ B, Ciro. Estadística Comercial, Editorial Norma, Bogotá, 1981.

D'OTTONE R, Horacio, Estadística Elemental, Editorial Cienes, Chile, 1971.

Centro de Administración, SENA. Unidades de Estadística, Ediciones anteriores.